

## Measuring Population Health Using Electronic Health Records: Exploring Biases and Representativeness in a Community Health Information Exchange

Brian E. Dixon<sup>a,b,c</sup>, P. Joseph Gibson<sup>a,d</sup>, Karen Frederickson Comer<sup>e</sup>, Marc Rosenman<sup>b,f</sup>

<sup>a</sup> Indiana University Fairbanks School of Public Health, Department of Epidemiology, Indianapolis, IN, USA

<sup>b</sup> Regenstrief Institute, Center for Biomedical Informatics, Indianapolis, IN, USA

<sup>c</sup> Veterans Health Administration, Health Services Research and Development Service, Indianapolis, IN, USA

<sup>d</sup> Marion County Public Health Department, Indianapolis, IN, USA

<sup>e</sup> Indiana University-Purdue University Indianapolis, Polis Center, Indianapolis, IN, USA

<sup>f</sup> Indiana University School of Medicine, Department of Pediatrics, Indianapolis, IN, USA

### Abstract

Assessment is a core function of public health. Comprehensive clinical data may enhance community health assessment by providing up-to-date, representative data for use in public health programs and policies, especially when combined with community-level data relevant to social determinants. In this study we examine routinely collected and geospatially-enhanced EHR data to assess population health at various levels of geographic granularity available from a regional health information exchange. We present preliminary findings and discuss important biases in EHR data. Future work is needed to develop methods for correcting for those biases to support routine epidemiology work of public health.

### Keywords:

Electronic Health Records; Health Information Exchange; Geographic Information Systems; Community Health Planning; Health Services Needs and Demand

### Introduction

Public health authorities monitor population health to identify burden of disease, manage health assets, establish policy, and evaluate interventions. This assessment usually relies on a limited set of information available through surveys, vital records, and paper-based disease reporting. Electronic health record (EHR) systems may provide more timely data for a larger portion of a population. Yet there exist a number of challenges to routine use of EHR data, including linking them to community data about social determinants of health. In this study, we sought to develop and evaluate neighborhood-level indicators of population health using EHR data integrated with a community information system (CIS).

### Methods

The Indiana Network for Patient Care (INPC), a large health information exchange with over 5 billion clinical observations from EHR systems, was geospatially enhanced and combined with social determinant data from SAVI, a community information system serving the same geographical region (1). We then assessed the prevalence of diseases of public health interest and calculated HEDIS-like clinical quality indicators. Using statistical methods we assessed the reliability and representativeness of these data to measure population health at various levels of geographic granularity

### Results

Rates of diabetes ranged from 1.5% to 16.07% with an average of 8.9% among neighborhoods spread across a metropolitan area. When examined at the census tract level, diabetes rates ranged from 1.5% to 12.83% with an average of 8.9% of the population for a given area.

We identified three biases in using EHR data. First, EHR data only represent those that seek health care. Second, linked EHR data are biased based how patient records were matched. The HIE uses a probabilistic technique, which can result in duplicate records. Third, the HIE proportionally contains more data from low income providers. We are exploring ways to adjust rates and correct these biases so they do not overestimate burden of disease and poor care quality in inner-city neighborhoods. We seek to compare census tract, neighborhood, and other geographic area measures with data from a recent population survey.

### Conclusion

EHR systems capture data about more people than do population surveys, but they have biases that affect their estimates of population health indicators such as disease prevalence or preventative screening rates. Future work is needed to develop methods for correcting for those biases to support routine epidemiology work of public health.

### Acknowledgements

This project was supported by the Robert Wood Johnson Foundation (No. 71271) by its National Coordinating Center for Public Health Services and Systems Research.

### References

- [1] Comer KF, Grannis S, Dixon BE, Bodenhamer DJ, Wiehe SE. Incorporating geospatial capacity within clinical data systems to address social determinants of health. Public Health Rep. 2011;126 Suppl 3:54-61.

### Address for correspondence

Brian E. Dixon, MPA, PhD, 410 W. 10<sup>th</sup> St, Suite 2000, Indianapolis, IN, 46202, USA [bedixon@regenstrief.org](mailto:bedixon@regenstrief.org)